

Allen Choi

allenyoungchoi@gmail.com | linkedin.com/in/allen-choi | San Luis Obispo, CA

EDUCATION

California Polytechnic State University

Master of Science in Statistics

Bachelor of Science in Statistics, Minor in Data Science

San Luis Obispo, CA

Expected June 2026

Expected June 2025

- **Cumulative GPA:** 3.74/4.00
- **Honors:** President's Honor List (2022-23, 2023-24), Mu Sigma Rho Statistics Honorary Society
- **Relevant Coursework:** Data Structures, Algorithms, Object Oriented Programming, Probability Theory, Estimation and Sampling Theory, Mathematical Statistics, Mathematical Foundations of Data Science

RELEVANT SKILLS

- **Languages:** R (tidyverse, ggplot2), Python (Pandas, PyTorch, tensorflow, scikit-learn), SQL, Java, SAS
- **Software:** Tableau, GitHub, Colab, Microsoft Office (Word, Excel, PowerPoint)

RELEVANT EXPERIENCE

Undergraduate Research Analyst

Sept 2024 - Present

California Polytechnic State University, San Luis Obispo

- **Processed** and **cleaned spatial and time series sensor data** using R (**dplyr**) to identify and resolve outliers while also using R (**leaflet**) to visualize rattlesnake temperature data across the US
- **Collaborated** with a cross-disciplinary team including professors from the biology department to explore factors affecting rattlesnake body temperatures

Undergraduate Data Science Researcher

June 2024 - August 2024

University of California, Irvine

- Participated in the **ISI-BUDS** research program which is funded by the National Institutes of Health (NIH)
- Utilized Seurat package in R for **clustering** and classification of distinct skin cell types across blister agent exposures
- Identified potential mechanisms of blister agent effect on skin through **differential gene expression analysis** and gene ontology (GO) enrichment analysis, sharing insights at a research symposium

Statistics Learning Tutor

Oct 2023 - Present

California Polytechnic State University, San Luis Obispo

- Facilitated tutoring as part of Cal Poly's Learning Support Center, supporting up to **10 different statistics courses**
- Provided **individualized tutoring** across **250+ sessions** in statistics fundamentals and software, helping students significantly improve understanding and boost performance in statistics courses

RELEVANT PROJECTS

Facial Recognition Model (Python - TensorFlow/PyTorch)

- Compared **deep learning neural networks** predicting ethnicity and gender using face images
- Developed and trained **Multi-Layer Perceptron (MLP)** and **Convolutional Neural Network (CNN)** models in TensorFlow/Keras, utilizing transfer learning with cutting-edge Residual Network (ResNet-50) in PyTorch
- Attempted to **address biases in AI predictions** through constant model evaluation and refinement, achieving notable improvements in accuracy

Cancer Prediction Model (Python - scikit-learn)

- Cleaned and preprocessed census data of **400k+** total participants' demographic information
- Utilized **scikit-learn** to construct and fine-tune a K-Nearest Neighbors, Random Forest, and an ensemble model to classify cancer history in people, optimizing hyperparameters to maximize recall scores
- Achieved **75% recall** for cancer prediction, demonstrating potential for early preventative detection

Cal Poly ASA DataFest 2024 (R/Python)

- **Collaborated in a team** for a 48-hour analysis competition sponsored by the American Statistical Association, utilizing R and Python to clean, analyze, and visualize complex datasets from a corporate sponsor
- **Applied machine learning techniques** to extract insights under 48 hours and delivered findings to a panel of judges

Customer-Restaurant Interface (Java/MySQL)

- Developed a **Java**-based customer application with **MySQL** database integration for easy data management, integrating purchase and checkout processes within a single application